The Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies

---

OMB CLEARANCE REQUEST

March 2005
Revised: October 2005

Note: Revisions to the original OMB Clearance Request for this project (OMB #: 1850-0801) are indicated in **bold** print throughout this document.

Prepared for:
Institute of Educational Sciences
United States Department of Education
Contract No. ED-04-CO-0111/001

Prepared by:
MDRC
16 East 34th Street
New York, NY 10016

TABLE OF CONTENTS

LIST OF EXHIBITS

INTRODUCTION

This document requests approval for the second phase of the data collection plan for The Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies (also known as the Enhanced Reading Opportunities study or ERO), a study of literacy interventions for ninth-grade students who read below grade level. In particular we are requesting OMB approval for an **additional set of data collection activities that will supplement the initial data collection (submitted to the OMB in March 2005 and approved in June (OMB #1850-0801)). The original submission requested approval for a student background questionnaire and a beginning-of-year teacher survey. The original submission document also included a discussion of the Group Reading Assessment and Diagnostic Evaluation (GRADE), a reading assessment being used as part of this study.** [1] **The March 2005 package also explained that it was the first part of a two-stage submission, and that there would be a supplementary request to follow later in 2005 for approval of subsequent data collection activities. Those activities are outlined in the original submission but not discussed in detail. The current submission includes a request for approval for the following:**

- **Three additional survey instruments to collect data from students and teachers.**

- **An additional year of data collection because the study has been expanded to include a second year of program implementation, involving a second cohort of students.**

We have included information about our additional data collection requests in the relevant sections of Part A: Justification and Part B: Collection of Information Employing Statistical Methods. Revisions to the original OMB clearance request for this project are indicated in bold print throughout the document.

The Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies addresses a pressing problem in education today – the need to improve the literacy of high school students. This evaluation will provide valuable data to inform the President's High School Initiative, which calls for all high school students to graduate with the knowledge and skills they need for good jobs and further education. Too many ninth grade students arrive at high school without the reading skills they need for high school coursework. Ninth grade is a pivotal year for youth—they must make the transition to more demanding subject material, prepare for academic proficiency exams often required for high school graduation, and manage a greater level of responsibility. Students who fall short in achieving these goals and are not successful during freshman year are more likely to fail to complete high school.

The present study makes an important contribution to the field of adolescent literacy, as rigorous scientific research on effective curricula for struggling adolescent readers is unavailable. The Department of Education's Institute for Education Sciences (IES), recognizing the dearth of

[1] The GRADE test form itself and the burden for its administration are excluded from the Paperwork Reduction Act requirements under 1320.3(h)(7).

experimental research about the impacts of adolescent literacy programs on reading and academic outcomes of struggling readers, has responded by commissioning this evaluation – the first large-scale randomized trial to study such impacts. Through a competitive process a team of researchers from not-for-profit research corporation MDRC and its subcontractor, American Institutes of Research (AIR), was selected to oversee the implementation and evaluation of two supplemental literacy interventions.

One strategy for providing support to struggling ninth grade readers at this crucial moment in their educational careers is to supplement their regular curriculum with a course that focuses on developing their literacy skills. So, instead of one of their electives or a study hall, they would attend a literacy class, while still carrying their required core courses and remaining on track for graduation. Based on what experts increasingly agree are necessary areas for effective literacy instruction, such a class would employ several key components: (1) motivation and engagement; (2) fluency, or the ability to read quickly, accurately, and with appropriate expression; (3) vocabulary, or word knowledge; (4) comprehension, or the ability to apply various strategies in order to extract meaning from texts; (5) phonics and phonemic awareness (for students who could still benefit from instruction in these areas); and (6) writing.

This evaluation will rigorously test two alternative literacy interventions with these components. A major feature of the evaluation is the use of an experimental design where random assignment is used to select a treatment group of students admitted to supplemental literacy classes and a control group of students who are not admitted and attend regular high school classes. Random assignment ensures that there are no systematic differences between the treatment and control group other than admission to the supplemental literacy classes. Thus systematic differences in the two groups in subsequent outcomes can be attributed to the effects of the treatment. The specific research questions intended to be addressed by this study are as follows:

1. Do specific research-based supplemental literacy interventions that create personalized and intensive instruction for striving ninth grade readers significantly improve reading proficiency, as measured by standardized reading achievement tests, by diagnostic reading inventories, and on tenth or eleventh grade state tests of English/Language Arts achievement?

2. What are the effects of the literacy interventions on other in-school outcomes such as attendance, persistence in school, and course-taking behavior? What is the nature of the relationship between improving reading skills and these other outcomes?

3. Do the interventions' impacts differ for students with different characteristics (e.g. varying levels of initial reading proficiency)? Which students benefit most from participation in the intervention?

OVERVIEW

As mentioned in the Introduction and described more fully in Section A: Justification, too many adolescents lack the literacy skills needed for success in school or at work. This study was developed by the Institute for Education Sciences (IES) in collaboration with the Office of Vocational and Adult Education (OVAE) to provide evidence on what kinds of interventions work or do not work to increase adolescent literacy. The study is incorporated into an OVAE initiative to support smaller learning communities (SLCs), which are thought to be particularly suitable environment for these literacy interventions.

OVAE held a special grant competition early in 2005 to award funds to support currently operating SLC programs that are willing to be part of a demonstration of two literacy programs. The OVAE grants include funds to pay for the supplemental literacy interventions. In addition, school districts awarded the SLC grants agreed, as part of the criteria for consideration, that their eligible schools would implement one of two supplemental literacy interventions (to which they are randomly assigned) and will participate in the experimental study – they will help identify students eligible for the interventions and accept that these students will be randomly assigned to treatment and control groups.

Study Approach

The Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies will be conducted by MDRC and its subcontractor American Institutes for Research (AIR). This section provides an overview of the study and the data to be collected. **We describe the approved baseline forms – the Student Background Questionnaire (with the Study Participation Agreement Form) and the Beginning-of-Year Teacher Survey. We include the description from the original OMB request of the Group Reading Assessment and Diagnostic Evaluation (GRADE), the reading assessment being used in this study. Additionally, we provide descriptions of the three additional data collection forms for which we are now seeking approval – the Student Follow-Up Questionnaire, and End of Year Teacher Surveys for both ERO and ELA teachers participating in this study.**

To address the research questions of the Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies, MDRC and AIR will: (1) randomly assign schools to one of two selected supplemental literacy programs; (2) randomly assign students who apply to the program in each school to a treatment group admitted into the program and a control group; (3) assure that the programs are implemented and document the implementation; (4) collect a variety of data to track student characteristics and outcomes; and (5) analyze the data and report the results of the analysis.

Random Assignment of High Schools. **Schools districts and schools were selected for this project through a special grant competition held by the U.S. Department of Education's Office of Vocational and Adult Education (OVAE) under its Smaller Learning Communities program. OVAE's selected 34 high schools in 10 school districts (4 in each of 7 districts and 2 in each of three districts) to participate. Each high school received funding**

to bolster Small Learning Communities (SLC) in the ninth grade - also known as Freshman Academies - and enable them to implement a supplemental literacy component. After selection, MDRC randomly assigned high schools within each district to one of the two supplemental literacy interventions: Xtreme Reading developed by KU-CRL and Reading Apprenticeship developed by WestEd. Within each district, half of its participating high schools were randomly assigned to one intervention and the other half were assigned to the other intervention. In all, 17 high schools are using each intervention. By randomly assigning schools to one of the two supplemental literacy interventions, the study will avoid the criticism that intervention developers were allowed to select "soft targets" (which would make the evaluation results inapplicable to schools less equipped to put the interventions in place) and create an implementation scenario more likely to be typical of schools in general.

Note that the study was initially designed to include 32 high schools. The Special OVAE grant competition resulted in 34 high schools being selected for the study. All burden estimates have been updated to reflect a sample of 34 high schools across 10 study districts.

Random Assignment of Students. As soon as districts and schools were notified of the OVAE funding decision, MDRC worked with the schools to identify all eighth-grade students registered for the high school whose seventh or eighth grade reading test scores make them eligible for the study, i.e., reading 2 to 4 years below grade level.

For each participating high school, approximately 150 students are to be given an informational brochure about the supplemental literacy classes, along with a set of forms (including a parental consent form) to be completed and returned. The pool for random assignment would include students identified as appropriate for the supplemental interventions on the basis of their pre-high school reading test scores who indicated interest and willingness to participate in the classes and in the study by returning the required forms. MDRC uses a computerized random assignment system to assign students to the program and control groups. Schools are to be notified of the results immediately, so that program group students can be scheduled for the special literacy courses and control group students could be placed in elective classes or a study hall. We will check that students have been assigned to the appropriate classes within the first weeks after random assignment and will continue to monitor their random assignment statuses over the course of the school year to prevent contamination of the experiment through control group students' participation in the supplemental classes.

The original Statement of Work for this study called for one cohort of students to be identified and randomly assigned to program and control groups. The study was designed to include approximately 100 students per high school (50 randomly assigned to a program group and 50 to be randomly assigned to a control group). This group of students (to be referred to as Cohort 1) includes students who were to be enrolled in the 9th grade during the 2005-06 school year. In September 2005, the Institute of Education Sciences approved a modification to the Statement of Work that included adding a second group of students to the study sample. This groups of students (to be referred to as Cohort 2) is to include students who are to be enrolled in the 9th grade during the 2006-07 school year. The study will conduct the same data collection activities for both cohorts of students.

Data Collection. The proposed study will draw on a variety of data sources to collect vital information on students, teachers, and classrooms. **The study received OMB approval for the Student Background Questionnaire and the Beginning-of-Year Teacher Survey. We are now asking for clearance on three additional data collection forms – the Student Follow-Up Questionnaire, the ERO End-of-Year Teacher Survey, and the ELA End-of-Year Teacher Survey.** The study uses:

A. Collections already approved

- a baseline Student Background Questionnaire to capture descriptive information about students just prior to random assignment; analysis of these data after random assignment will ensure that the assignment worked as intended
- reading assessments to measure both students' baseline reading skills and the skills they acquire over the course of the school year (excluded from this process per 1320.3(h)(7))
- a survey of teachers responsible for teaching one of the two reading curricula at baseline to provide descriptive data on their background characteristics, preparation, perceptions of curricular reform efforts, and participation in professional development activities
- school administrative records of measures of academic progress – test scores, grades, courses attempted and passed and credits earned toward graduation.

B. Follow-up data collections whose clearance is sought with this revised package

- **a Student Follow-Up Questionnaire to gather descriptive data about students after participation in the treatment and control programs to assess treatment program outcomes**
- **a follow-up survey of treatment teachers at the end of the school year to measure their impressions of the interventions and the implementation of the reading curricula**
- **a survey of English/language arts (ELA) teachers at the end of the school year to gather information which will allow us to understand better the literacy education contexts in which the programs are being implemented and where we expect to see site variation.**

Analyzing Data. The analysis will focus on estimating the impacts of the two literacy interventions on student reading ability and progress in school. Analysis will be done on a number of levels. The first level of analysis will compare all students in the treatment groups to all students in the control groups. The second level will estimate the impacts on students for each of the two interventions and for effects on subgroups of students. The third level will explore sources of impact and variations in the sizes of impacts.

This submission provides an overview of all aspects of the planned data collection. It provides details on the Student Background **and Follow-Up** Questionnaires, and the GRADE Reading Assessment, both of which provide important data for the analyses of student outcomes.

It also provides details on the Beginning-of-Year **and the End-of-Year** Teacher Surveys, important pieces of our study of the implementation of the intervention.

PART A: JUSTIFICATION

1. Circumstances That Make Collection of Data Necessary

Data from the 2002 NAEP indicate that 38 percent of the 12th grade students tested at the "Basic" level, where "Basic" denotes only *partial mastery* of the knowledge and skills deemed fundamental for proficient work at the grade level, while 26 percent of the students tested at the "Below Basic" level (Grigg, Daane, Jin, & Campbell, 2003). Only 36 percent of the students registered scores at the "Proficient" or "Advanced" levels, demonstrating competency over challenging subject matter.

The NAEP results are especially troubling on many counts. First, they pertain only to students who were still enrolled in school – and available to be tested – in 12th grade. They therefore say nothing about the literacy skills of the large proportions of students – as high as 50 percent in some urban high schools – who drop out at an earlier point. Since dropouts are likely to be weaker students than those who persist, we can only assume that their reading scores would have been even lower than those registered by the test-takers.

Second, the findings point to sizable and persistent gaps in performance between subgroups defined by race or ethnicity, place of residence, economic class, and gender. Children from European-American households scored higher than children from African-American families, Latino families, and families of other backgrounds where the first language or language spoken at home is not English; children in urban settings scored higher than their rural counterparts; middle-class children scored higher than working-class children; and girls scored higher than boys.

Third, the results indicate a decline in performance since 1992, when 59 percent of the 12th-grade students scored at the Below Basic or Basic levels (compared with 64 percent in 2002).[2] And finally, this decline has occurred at the same time that literacy requirements in secondary schools and in the workplace are on the rise. Not only is there more for high school students to read and learn, there are also more testing hurdles en route to graduation. As of 2000, 28 states had enacted laws or had legislation in the works to require that high school students pass exit examinations assessing sophisticated reading skills in order to graduate (Reardon & Galindo, 2002).[3] In addition, the influx of technology in schools and in homes raises the level of

[2] It is likely that demographic shifts account for some of this change and complicate the ability of schools to effect literacy growth for adolescents. More and more English language learners are entering America's public schools, and this trend will continue for some time: The National Center for Educational Statistics predicts that by 2020 students from cultural and language minority groups will be in the majority. A discussion of the controversy about how best to promote the development of English language learning is beyond the scope of our proposal; here we wish only to point out that the changing composition of the student population has major implications for literacy instruction.

[3] It is worth noting that high school graduation does not ensure strong reading skills. Students entering college are expected to have a reading proficiency above 12th-grade level, as determined by a traditional reading comprehension test. But data from the National Center for Education Statistics (2004) indicate that 11 percent of freshmen entering degree-granting institutions that offered remedial coursework took remedial classes in reading,

reading skills needed by adolescents. Students must master these expanding and evolving communication systems in order to use them effectively in the working world (Rycik & Irvin, 2001).

The special demands of ninth grade. Research by MDRC and others shows that ninth grade is a "make-or-break" point for many high school students. Students who lack sufficient literacy and mathematical skills may fail to accumulate enough course credits for promotion, and failure to be promoted on time to tenth grade has been shown to be a strong predictor of subsequent dropout (Kemple and Herlihy, 2004). Thus, interventions that help students acquire the requisite literacy skills early on can help forestall a high school trajectory marked by persistent failure.

Studies confirm that the reading challenges that students face increase markedly over their academic careers. As students progress through the primary grades to the middle grades and then high school, they read increasingly complex textbooks, supplementary materials, and electronic text. In particular, the reading requirements of ninth grade represent a new and giant leap for entering freshmen, who face an increase in the amount of reading that is required in their courses, textbooks that are thicker and more intimidating than in previous grades, and a vocabulary load in content area instruction that can be overwhelming. It is not surprising that even students who appear to be strong readers in middle school may falter as they approach the demands of high school reading (Caldwell & Leslie, 2003/2004). Struggling readers, who may harbor real interest in their academic subjects but lack confidence in their ability to improve their reading, may feel uncomfortable in school, may increasingly avoid challenging reading materials, and may try to avoid situations in which their poor reading skills will be exposed (Guthrie, 2002; Guthrie & Alvermann, 1999; Wigfield, 2004).

The unique contributions of this study. The study will make a substantial contribution to our understanding of adolescent literacy. While many studies have led to a greater understanding of the reading skills that young children need and how best to teach these skills (National Reading Panel, 2002; National Research Council, 1998), the literature on the reading strengths and deficits of students who have reached adolescence is still relatively sparse.[4] Although enough is known to delineate the skills deficits and other issues that literacy interventions directed toward adolescents must address, this study will add to provide evidence on whether two specific interventions that embody the "best practices" actually can improve students' literacy skills substantially and in a relatively short period of time.

The use of a random assignment design to study the effects of the interventions helps ensure that, if the design is carefully executed and the interventions are well implemented, the study will yield the strongest, most reliable evidence possible on which to base policy and practice. Partly responding to the No Child Left Behind requirements and to the institution of

---

and 14 percent did so in writing. In public two-year colleges, the proportions were higher: 20 percent in reading and 23 percent in writing.

[4] The Alliance for Excellent Education (2004; Kamil, 2003), the International Reading Association (Moore, Bean, Birdyshaw, & Rychik, 1999), and the National Reading Conference (Alvermann, 2001) have all published discussions relating to policies for adolescent literacy instruction.

high-stakes state tests, high schools are now beginning to implement initiatives to address the needs of students with poor reading skills. The results of this study will help policy makers and administrators at all the federal, state, district, and school levels make the best choices about how to help the students they serve. Improving adolescent literacy is a key priority for ED; as this study is an important contribution to both the High School Initiative and to a foundation of knowledge for the President's Striving Readers Program.

Two cohorts of students. Given concerns that were raised early in the project (for example, by OVAE's Assistant Secretary Susan Sclafani and by the evaluation's technical advisory panel) that collecting data on only one year of implementation—during a period when the teachers are only just becoming familiar with the intervention approach and materials—may not be a good test of the effectiveness of the programs, OVAE expanded the special grants to districts and schools to require a second year of literacy program implementation, and has asked IES to expand the evaluation design accordingly. Evaluating a second year of the program not only allows ED to assess the programs when they are more fully implemented but also allows us to determine whether there really are differences in results between the first and second years of an intervention. Thus the additional cohort of students critically improves what we can learn from this study, positioned to contribute unique knowledge to the field of adolescent literacy. These decisions to expand the study to a second cohort were made after the original forms clearance package was submitted. They are presented in more detail in this revised package as we indicated we would do in our response to a question from the OMB regarding our original submission about why a one-year as opposed to a multi-year study was proposed.

Each of the data sources described below is an integral part of the evaluation. The Student Background **and Follow-Up** Questionnaires, and the GRADE will be used to describe the students in the study, to define subgroups to see if the interventions work better for some students than others, and to provide covariates to increase the precision of the analyses. The Beginning **and End** of Year Teacher Surveys provide us with necessary data about aspects of the implementation of the programs, such as who is implementing them and their impressions of the programs and the strength (or weakness) of their implementation. This information allows us to understand the context in which these programs are implemented, to interpret our impact findings more carefully, and to understand how these programs might be scaled up and adapted to other schools and districts.

2. Purposes and Uses of the Data

This section describes the data to be collected as part of this study. As noted earlier, OMB has approved our use of baseline data collection forms; thus, at this time, we are only asking for approval of the supplemental forms to be used – the Student Follow-Up Questionnaire and the End-of-Year Teacher Surveys. Descriptions of the other data sources, the GRADE in particular, are provided to put the baseline and follow-up data sources in context with the overall plan. The questionnaires and the literacy assessment are important research tools that will provide variables used in the analysis the outcomes. **The**

instruments described will be administered during both of the two school years in which the programs under study will be implemented.

Baseline Data

Student Background Questionnaire. This questionnaire, adapted from the 2003 NAEP Student Reading Background Questionnaire for eighth graders, is necessary to gather data for this study that would be unavailable otherwise. It provides us with demographic, educational, and literacy background information about the study participants. These data are essential for describing the study population in ways that inform our understanding of the generalizability of the study results and make it clear to readers of study reports who the participants in the study were.

These data are also important for identifying subgroups and covariates for analyses of the outcome measures. They allow us to answer questions about whether there are differential outcomes for students with different literacy or educational resources at home, of different racial backgrounds, or with different reading attitudes and prior reading behaviors. Table 1 describes how each question in the survey will be used.

Table 1: Student Background Questionnaire – Item-Level Justification

Item	*Topic*	*Purpose*
1	Racial background	To investigate whether students of different racial/ethnic backgrounds benefit differently from the intervention. Although racial and ethnic background data is often available through school records, school districts often do not use the same categories (unlike gender data, for example). Using the questionnaire for this data provides consistency across sites.
2-7	Literacy and educational resources at home	To investigate whether student reading and educational outcomes vary with the amount and types of literacy and educational resources students have in their homes.
8	Required school reading	To provide information about how much school reading the students have been doing. Do students who required to do different amounts of reading for school benefit differentially from the interventions?
9	Student attendance	To investigate if students have had past attendance issues, and if so, whether the interventions have different impacts on students with attendance problems. Ideally we will be able to get prior attendance information from school records, but this item provides us with back-up in case some schools have purged their records.
10-11	Mother's and father's education	To assess potential differences in program impact by this measure of family socio-economic status and educational capital in the home.
12	Non-English language at home	To address whether there are differential impacts of the interventions based on how often English is spoken at home.
13	Literacy attitudes	To determine if students who already see value in and enjoy reading and writing get more out of the intervention than students who enter with more negative literacy attitudes. Even if there is little variation within the study sample on this item, it will contribute to a literacy profile necessary in a description of the study sample.
14	Non-school literacy behaviors	To determine if students who are more engaged in literacy activities get more out of the intervention than students who are less engaged. Again, as stated above, even if there is little variation within the study sample on this item, it will contribute to a literacy profile necessary in a description of the study sample.
15-17	Non-school reading behavior by type of material	To assess students' the fiction, nonfiction, and media literacy experiences. Does the program have different impacts for students more or less engaged in different literacy categories?

Reading Assessment Test. Student reading achievement is a critical data element for the impact study. Pre-intervention achievement will be used in a number of ways: as a descriptor of the students in the study; as an important attribute on which to define subgroups for the program participation and impact analyses; as a baseline characteristic that can be used as a covariate in the analysis models to refine the standard errors of the impact estimates; and as a benchmark against which to calculate growth in both specific literacy skills and overall literacy. Finally, the fall assessment can serve as a check on the appropriateness of students in the study sample for the supplemental literacy interventions (although once random assignment has taken place, students must remain in the research sample under their designated research status).

We are using Form H of the *Group Reading Assessment and Diagnostic Evaluation*[5] as the diagnostic reading test to be administered to **both cohorts of** students at the beginning and end of ninth grade **during the two years that the literacy interventions will be used for this study**. This test, also known as GRADE, has been developed by American Guidance Service and is a norm-referenced battery of items that can provide detailed information about students' reading. The test, which provides among other data grade equivalency scores, will confirm or disconfirm student eligibility for the study and also provide valuable profiles of their strengths and weaknesses for use by intervention tests.

GRADE Form H yields valid and reliable information for low-, middle-, and high- readers at the high school level. This form of GRADE has been specifically developed for use with high school students. There are two forms of the test, so pre- and post-testing is possible. The test is untimed; the maximum administration time is estimated to be 90 minutes if all subtests are administered in the same session.[6] Because subtests can be administered separately, the test can be divided across two or more sessions to accommodate schools' class schedules.

Constructs Measured on the Diagnostic Reading Test

GRADE Form H assesses vocabulary, reading comprehension, and listening comprehension. The items on the test sample the domain of reading widely, with items targeting the following:

> Vocabulary: Provides measures of students' decoding and understanding through assessment of vocabulary knowledge with the context of short phrases or sentences.
> Sentence Comprehension: Provides measures of students' ability to comprehend sentences as whole or complete thoughts; this information supplements information resulting from the subtest that assesses comprehension of longer prose selections.
> Passage Comprehension: Provides measures of students' ability to read and understand passages of one or more paragraphs; the passages, which are designed to be age-appropriate and engaging, cover a full range of content areas and genres appropriate to high school students (e.g., literature, history, science, etc.).

[5] *Group Reading Assessment Diagnostic Evaluation.* (2001). Circle Pines, MN: American Guidance Service, Inc.
[6] With the exception of one test, the other tests that were considered were longer than the GRADE. The shorter test, the Gates-MacGinitie, 4th edition, takes approximately 55 minutes, which is beyond the usual length of a single class period. It provides information on vocabulary and comprehension without the detailed profile that the GRADE can provide.

Metacognition: Provides measures of students' use of diverse reading comprehension skills, including questioning, predicting, clarifying, and summarizing.
Listening Comprehension: Provides measures of students' ability to comprehend oral messages that require high levels of vocabulary, understanding of non-literal language, and the ability to make inferences.

Advantages of GRADE for This Study

The GRADE is only one of the many diagnostic tests available for use with ninth grade students. It does, however, provide the most detailed information about students' performance that can be obtained from a group administered test. Data from state reading tests are often very limited: students receive global scores that do not provide detailed information about what they know and are able to do when they read diverse kinds of texts. The GRADE, with its many subtests and its use of passages representing different genres and content areas, will provide a profile of students' strengths and weaknesses that can be helpful as the intervention begins. Further, this profile of reading abilities will facilitate fine-grained analyses of changes in students' reading performance at the end of the intervention year. The GRADE subtests and their resulting subscores provide the best reflection of any of the available tests of the dimensions of reading (e.g., comprehension, vocabulary, metacognition) that measure student performance on the components of the high quality supplementary literacy interventions that will be used in the study.

Data resulting from GRADE administrations include standard scores, percentile ranks, normal curve equivalents, stanines, grade equivalent scores, and Growth Scale Value (GSV). These data, coupled with the detailed profiles of students on the various GRADE subtests, will allow for in-depth analyses of changes in students' reading performance from the beginning to the end of the study year.

Implementation

Beginning-of-Year Teacher Questionnaire. This survey will be administered **annually** to teachers participating in the intervention, i.e. teachers who are responsible for implementing the supplemental curricula with the approximately 50 struggling readers in the treatment group. This questionnaire is necessary to gather data that would otherwise be unavailable. It includes items that are adapted from the teacher surveys from the National Longitudinal Evaluation of Comprehensive School Reform, the Evaluation of the National School District and Network Grants Program (Gates Foundation), the Alabama Reading Initiative, the Teacher Belief Survey (Benjamin, 2003), and the Survey of Instructional Practices and Content for English Language Arts and Reading (Survey of Enacted Curriculum, 2004). Respondents will be the intervention teachers, along with back-up teachers who might step in to teach the intervention during the school year if there is any teacher attrition.

The questionnaire has two purposes. First, it is designed to provide us with demographic and educational/training background information about the teachers who will deliver the intervention. Like the Student Background Questionnaire, these data are essential for describing the study population in ways that inform our understanding of the generalizability of the study

results and make it clear who the instructors of the intervention conditions are. They provide information about the contexts in which each of the two supplemental literacy programs are being implemented.

The second purpose of the questionnaire is to measure teacher characteristics, attitudes, and perceptions that (a) we think are likely to influence that person's capacity to run a supplemental reading program and will thus contribute to our understanding of program implementation at the participating schools; or (b) we think could be influenced by their involvement in the professional development and implementation of the literacy programs. Information on teacher perceptions and acceptance of the intervention is crucial for generalizing study results, specifically as it relates to future scale-up of this type of literacy intervention.

Because portions of the questionnaire will be administered again at the end of the year, these data will allow us to measure change in intervention teachers' perceptions from the beginning to the end of the year. Since the focus of the research design for this project is on impacts on student achievement, we will not be able to draw definitive causal conclusions that changes we may observe in teachers' perceptions were the result of their participation in implementing the intervention. However, learning whether teachers felt more or less positive about the supplemental literacy program they implemented by the end of the year is valuable descriptive, contextual data.

Table 2 (below) describes how each question in the survey will be used.

Table 2: Beginning-of-Year Teacher Questionnaire – Item-Level Justification

Item	*Topic*	*Purpose*
1	Gender	To investigate whether teachers of either gender implement or perceive the intervention differently.
2-5	Certification/education/ experience teaching	To investigate whether teachers of different experience and educational backgrounds implement or perceive the intervention differently.
6	Racial background	To investigate whether teachers of different racial/ethnic backgrounds implement or perceive the intervention differently.
7-8	Literacy/reading professional development experience	To assess whether the amount or focus of teachers' previous professional development experience is associated with implementation or perception of the intervention.
9	Perception of school climate/school change	To assess whether teachers' perception of school climate or school change is associated with implementation or perception of the intervention. To establish a baseline against which to measure change in intervention teachers' perceptions from the beginning to the end of the year.
10	Perception of student learning	To measure teachers' attitudes regarding student learning and whether those attitudes relate to implementation quality. To establish a baseline against which to measure change in these attitudes from the beginning to the end of the year.
11	Perception of intervention training	To assess teachers' perceptions of the initial intervention training and whether they are associated with implementation quality.
12-14	Perception/expectations for intervention	To measure teachers' initial perceptions of the intervention and whether these perceptions are associated with implementation quality.

<u>End-of-Year Teacher Surveys.</u> An End of Year survey will be administered to teachers of the supplemental literacy classes (referred to as the ERO Teacher Survey) and to teachers who provide English/language arts (ELA) instruction to ninth graders in the study schools (referred to as the ELA Teacher Survey). The survey will be administered annually near the end of the academic years in which this study is conducted. It will allow us to determine in more detail the nature of the service contrast between the additional literacy support the program students receive from the intervention compared to what they and the controls receive as part of their regular ELA curriculum. A second goal of this survey will be to collect data about the intervention teachers' perceptions of the

intervention and its implementation after having taught the supplemental literacy course for the years in which the study is conducted. There will be two versions of this survey, one specific to the intervention teachers, and one specific to English/language arts teachers, and these are both submitted for approval as part of this revision.

ERO teacher sample:

The survey of teachers responsible for teaching one of the two reading curricula at the end of the year will provide us descriptive data on their perception of the implementation of the reading curricula, instructional practices that occurred during the year, and their overall impression of the interventions and support for the intervention.

ELA teacher sample:

The survey of ELA teachers will allow us to determine in more detail the nature of the instruction that intervention and control student receive as part of their regular ELA, particularly as it pertains to literacy. The value-added of the literacy interventions may be lessened by strong ELA literacy instruction, or the programs might be strengthened by complementary activities going on in ELA classrooms. Understanding ELA instruction, particularly that focused on literacy activities, and the background and training experience of ELA teachers are thus valuable for describing the conditions under which the programs were implemented and may help in the interpretation of impact findings.

Table 3: ERO and ELA Teacher End-of-Year Surveys – Item-Level Justification

ERO teacher items	*ELA teacher items*	*Topic*	*Purpose*
	1-6	**Background and education; experience teaching**	**To determine in more detail the nature of the service contrast between the instruction that intervention and control student receive as part of their regular ELA and the additional reading support the intervention students receive from the program. [a]**
	7-8	**Literacy/reading professional development experience**	**To determine in more detail the nature of the service contrast between the instruction that intervention and control student receive as part of their regular ELA and the additional reading support the intervention students receive from the program.[a]**
1-2	**9-10**	**Instructional support**	**To determine the nature of the instructional support that intervention teachers may receive that is above and beyond the support experiences by ELA teachers.**
4	**11-12**	**Instructional practices**	**To determine the nature of the contrast between the instruction that intervention and control student receive as part of their regular ELA and the additional reading support the intervention students receive from the program.**
6		**Perception of school climate**	**To assess whether teachers' perception of school climate or school change is associated with implementation of the intervention. Change can be measured for intervention teachers' perceptions from the beginning to the end of the year. [a]**
7		**Perception of student learning**	**To assess whether teachers' perception of student learning is associated with implementation of the intervention. Change can be measured for intervention teachers' perceptions from the beginning to the end of the year. [a]**
3, 5, 8-9		**Teacher and student practices related to intervention**	**To assess whether teachers' perceptions of teacher and student practices related to the intervention are associated with implementation quality.**
10-11		**Perception for intervention**	**To measure whether teachers' perceptions of the intervention are associated with implementation quality. Change can be measured for intervention teachers' perceptions from the beginning to the end of the year. [a]**

[a] Items given to intervention teachers on baseline survey.

Outcome Measures

Data on student outcomes will come from three sources: literacy assessments administered in the spring of students' ninth-grade year and school administrative records following the students' tenth grade year (as approved in the original submission), and also from the Student Follow-Up Questionnaire (for which we are requesting approval now). In addition, because the study has been extended one year, we are requesting to collect these data from a second cohort of students, and also to collect school administrative records for an additional year – after the first cohort completes eleventh grade and the second cohort completes tenth grade.

<u>Reading Assessment Test.</u> This follow-up assessment will occur for two cohorts of students instead of one – in the spring of 2006 for Cohort 1 and in the spring of 2007 for Cohort 2.

<u>School records data.</u> Because an additional year of data collection has been added to this study, we will be collecting data through 2008. This collection effort will include data about the eighth- through eleventh-grade years of the first cohort and the eighth- through tenth-grade years of the second cohort. Data sought from all students will include student characteristics and measures of academic progress – test scores, grades, absences, courses attempted and passed, and credits earned toward graduation.

<u>Student Follow-Up Questionnaire.</u> The Student Follow-Up Questionnaire will be given to both cohorts of students at the end of their ninth-grade year. It will be used to measure outcomes that the treatment programs could be expected to affect around educational and literacy behaviors and attitudes. The developers of both programs expect that their programs will affect not only students' literacy skills and academic outcomes, but also their attitudes and behaviors more generally as related to school and literacy activities. This survey will be administered to treatment and control students and is submitted for approval as part of this revision. Table 4 (below) presents item-level justifications.

Table 4: Student Follow-Up Questionnaire– Item-Level Justification

Item	*Topic*	*Purpose*
1-2	Attitudes about going to school and school behaviors	To investigate whether the interventions affected students' attitudes toward school and their behavior. Constructs reflect intrinsic and extrinsic motivations for attending school. Behavior items provide common measures across all sites. School administrative records are likely to vary in the detail and types of disciplinary and student behavior information they contain.
3	Expectations about future education	To determine whether the interventions had an impact on student expectations of future education. Both intervention programs address students' expectations for the future.
4-8	Literacy attitudes and out-of-school literacy behaviors	Several sub-items of these questions are also asked in the baseline instrument. They will enable an investigation of changes in the students' literacy attitudes and behaviors as a result of their participation in the supplemental literacy classes.
9-20	Literacy supports in and out of school	These items help us understand the kinds of literacy supports students may have in and out of school that could effect literacy gains in both treatment and control groups. These data will be crucial for understanding both the treatment contrast between program and control students and the context of literacy support in which the programs were implemented.
21-22	In-school literacy behaviors	To determine if there is an impact on the facility with which students read school materials and how often they do reading for school.
23-28	Work and reading for core subjects	These items offer the opportunity to explore more specifically what impacts these literacy interventions may have on students' approaches to their core subject area classes.
29-31	Attitudes about ERO classes	These items would be asked only of the students in the program group to investigate their perceptions of their experiences in and the usefulness of their Enhanced Reading Opportunities classes.

Privacy Issues

The Department of Education may disclose information contained in these data under the routine uses listed in the system of records without the consent of the individual if the disclosure is compatible with the purposes for which the record was collected.

Freedom of Information Act (FOIA) Advice Disclosure. The Department may disclose records to the Department of Justice and the Office of Management and Budget if the Department concludes that disclosure is desirable or necessary in determining whether particular records are required to be disclosed under the FOIA.

Contract Disclosure. If the Department contracts with an entity for the purposes of performing any function that requires disclosure of records in this system to employees of the contractor, the Department may disclose the records to those employees. Before entering into such a contract, the Department shall require the contractor to maintain Privacy Act safeguards as required under 5 U.S.C. 552a(m) with respect to the records in the system.

Research Disclosure. The Department may disclose records to a researcher if an appropriate official of the Department determines that the individual or organization to which the disclosure would be made is qualified to carry out specific research related to functions or purposes of this system of records. The official may disclose records from this system of records to that researcher solely for the purpose of carrying out that research related to the functions or purposes of this system of records. The researcher shall be required to maintain Privacy Act safeguards with respect to the disclosed records.

We expect that this collection will have minimal impact on privacy. Results will never be disaggregated and reported in such a way that individuals can be identified. Only persons conducting this study and maintaining its records will have access to the records collected that contain individually identifying information. The first step in working with the collected data is to encrypt individual identifiers so that analyses of the data are conducted on anonymous data.

3. Use of Improved Technology to Reduce Burden

Whenever possible we will use information technologies to maximize the efficiency and completeness of the information gathered for this evaluation and to minimize the burden on respondents. In particular, whenever possible we will collect data from existing electronic school administrative records.

4. Efforts to Avoid Duplication

The Student Background **and Follow-Up** Questionnaires were designed to include only questions for which data could not be obtained consistently across sites through other methods. The same is true of the Beginning **and End** of Year Teacher Surveys. We intend to administer our own literacy assessment, the GRADE, even though student results on state eighth grade assessments represent existing data about students' reading level, and these data will be used to determine the universe of those who read 2 to 4 years below grade level, the key eligibility criterion for participating in this study. However, these state assessments vary in what they measure, how closely they can measure it, and how well they can measure student reading achievement. Thus, administering our own assessment is necessary because it provides site-to-site consistency. It also gives us a level of detail about students' reading abilities that state assessments do not provide.

5. Efforts to Minimize Burden on Small Businesses or Other Entities

No small business or entities will be involved as respondents.

6. Consequences if the Information Is Not Collected or Is Collected Less Frequently

The Student Background Questionnaire will be administered at the beginning of the 2005-2006 school year **for Cohort 1 and at the beginning 2006-2007 for Cohort 2.** If it is not administered and the data not collected, both the descriptive and impact analyses will not be adequate. This lack of detail in the description of the study population will limit both our interpretations of the results in the context of this study and our ability to generalize about the results to other situations. It will also not be possible to do valuable subgroup analysis and the statistical analyses will be less precise because we will lack important control variables.

The Student Follow-Up Questionnaire will be given to both cohorts of students at the end of their ninth grade year. If it is not administered, we will not have data about student outcomes that are valuable to our understanding of the students' perceptions of their own literacy abilities, behaviors, and attitudes about reading and school in general, providing us with information about how literacy programs affect the self-efficacy of students. Further, some of these variables will be used as outcome measures helping to identify the extent to which the ERO classes increase the amount of student reading or writing outside of school. Collecting these data will also provide us with a better understanding as to whether or not non-cognitive factors such as attitudes and behaviors are related to student achievement. Both programs being tested in this study are expected to have impacts on these types of outcomes, and we believe that collecting data about them will aid the interpretation of the results of the analyses of academic outcomes and improve our understanding of the nature of these programs. Also, literacy research suggests the importance of learning about non-academic literacy outcomes (National Assessment Governing Board, 2003; Biancarosa & Snow, 2004).

The Beginning of Year Teacher Survey will be administered **annually**, at the summer professional development trainings just prior to the start of the 2005-2006 **and 2006-2007** school years. If it is not administered and the data not collected, we will not have data about the teachers charged with implementing (i.e., teaching) the supplemental literacy programs. These data are essential for us to understand how these teachers, based on background characteristics, training, and their perceptions of literacy education and reform efforts, affect the implementation of the programs. These data provide context for our study that facilitates our ability to understand how these programs might work in other settings with other personnel.

The two End of Year Teacher Surveys will be given in the spring of 2006 and 2007. If these data are not collected, we will not have information to consider the roles of the teachers in the student achievement outcomes of the study. These data will also serve an important descriptive purpose, allowing us to understand the context in which the interventions are mounted. These data will provide us with information useful for investigating issues associated with implementation and interpreting the results of the impact analyses.

The GRADE will be administered twice annually with both cohorts of ninth-graders participating in this study: as a baseline measure before or at the start of the school year

and as an outcome measure at the end of the school year. The GRADE will be administered to students in Cohort 1 at the start and end of the 2005-2006 school year and to students in Cohort 2 at the start and end of the 2006-2007. The baseline administration of the GRADE will provide the same measure of reading ability across all schools and will serve as an important control variable and in determining subgroups for analysis. The spring GRADE is one of the primary outcome measures and if it is not collected the study would be seriously curtailed

7. Special Circumstances Requiring Collection of Information in a Manner Inconsistent with Section 1320.5(d)(2) of the Federal Regulations

No special circumstances apply to this study.

8. Federal Register Comments and Persons Consulted Outside of the Agency

A 60 day notice was published in the Federal Registry on January 4, 2005 with an end date of March 7, 2005 to provide for public comment.

The following individuals were consulted in the development of these materials:

Dr. James Kemple, MDRC

Dr. Terry Salinger, American Institutes for Research

Dr. Jason Snipes, MDRC

9. Payment to Respondents

No payments will be given to student, parents, or teachers for the Student Background Questionnaire, **Student Follow-Up Questionnaire**, Study Participation Agreement Form, the Beginning of Year Teacher Survey, or **the End of Year Teacher Surveys**.

10. Assurance of Confidentiality Provided to Respondents

All data collection activities will be conducted in full compliance with The Department of Education regulations to maintain the confidentiality of data obtained on private persons and to protect the rights and welfare of human research subjects as contained in the Department of Education regulations. These activities will also be conducted in compliance with other federal regulations in particular with The Privacy Act of 1974, P.L. 93-579, 5 USC 552 a; the "Buckley Amendment," Family Educational and Privacy Act of 1974, 20 USC 1232 g; The Freedom of Information Act, 5 USC 522; and related regulations, including but not limited to: 41 CFR Part 1-1 and 45 CFR Part 5b and, as appropriate, the Federal common rule or ED's final regulations on the protection of human research participants.

The organizations that are part of the research team will follow procedures for assuring and maintaining confidentiality that are consistent with the provisions of the Privacy Act. The following safeguards are routinely employed to carry out confidentiality assurances:

- All staff members – at MDRC and AIR – sign an agreement to abide by the corporate policies on data security and confidentiality. This agreement affirms each individual's understanding of the importance of maintaining data security and confidentiality and abiding by the management and technical procedures that implement these policies.

- All data, both paper files and computerized files, are kept in secure areas. Paper files are stored in locked storage areas with limited access on a need-to-know basis. Computerized files are managed via password control systems to restrict access as well as to physically secure the source files.

- Merged data sources have identification data stripped from the individual records or are encoded to preclude overt identification of individuals.

- All reports, tables, and printed materials are limited to presentation of aggregate numbers.

- Compilations of individualized data are not provided to participating agencies.

- Confidentiality agreements are executed with any participating research subcontractors and consultants who must obtain access to detailed data files.

An explicit statement describing the project, the data collection and confidentiality will be sent to the student and needs to be signed by student and a parent for the student to become part of the evaluation. The proposed consent form – Study Participation Agreement Form – is attached.

A Privacy Impact Assessment has been conducted (see the discussion of Privacy Issues in section A.2. above) and the Privacy Act System of Notice is currently being developed.

We prepared a System of Records (SOR) and notice was published in the Federal Register on June 24, 2005.

11. Justification for Questions of a Sensitive Nature

No questions of a sensitive nature are included in the Student Background **or Follow-Up** Questionnaires, or the Beginning **or End** of Year Teacher Surveys.

12. Estimate of Information Collection Burden

Table 5 below provides annual burden estimates for students for the Student Background Questionnaire and the Beginning of Year Teacher Survey which were updated from the original OMB submission, reflecting the participation of two additional schools in the demonstration. Because these are *annual* estimates, they were not adjusted based on the addition of a second cohort of students and teachers. However, this estimated burden is now applicable to two years of data collection rather than only one. The new version of Table 5 also includes annual burden estimates for the Student Follow-Up Questionnaire, the End-of-Year Teacher Survey, and ELA Teacher Survey. This total annual burden for data collection is estimated at 1,998 hours including both revisions to the initial estimates and the addition of the new data collection instruments.

Table 5: Annual Burden Estimates by Task

Instrument	*Estimated Number of Respondents*	*Average Time Per Response (Hours)*	*Total Respondent Time (Hours)*
Student Background Questionnaire and Consent Form: 70% response rate (based on average of 150 potentially eligible students per school)	**3,570**	**0.25**	**892.5**
Beginning-of-Year ERO Teacher Survey: 100% response rate	**34**	**0.25**	**8.5**
Student Follow-Up Questionnaire: 90% response rate (average of 105 students per school)	**3,213**	**0.33**	**1,060**
ELA Teacher Survey: 75% response rate (based on 4 ELA teachers per school)	**102**	**0.25**	**25.5**
End-of-Year ERO Teacher Survey: 100% response rate	**34**	**0.33**	**11**
TOTAL:	**6,953**	**0.29**	**1,998**

13. Estimate of Total Annual Cost Burden to Respondents

There are no direct costs to participants.

14. Estimates of Annualized Costs

The estimated cost to the federal government of conducting the Evaluation of the Impact of Supplemental Literacy Interventions in Freshman Academies is based on the government's contracted cost of the data collection and related study activities along with personnel cost of government employees involved in oversight and/or analysis. For all data collection activities for which OMB approval has been requested, *which now include two cohorts of students and follow-up surveys for students and teachers*, the overall cost to the government is $2,378,466 (an average of $1,189,233 per cohort).

This cost includes:

- **$1,090,838 for activities surrounding random assignment, including obtaining parental consent and collecting baseline intake data;**
- **$220,306 to conduct the teacher surveys;**
- **$436,145 for collection of administrative records;**
- **$631,177 to conduct the literacy assessment: GRADE.**

15. Change in Annual Reporting Burden

The total annual data collection burden presented in the original OMB Clearance Request submitted for this project was 975 hours for an estimated 3901 respondents and an average of 0.25 hours per respondent. Based on revisions to the initial estimates, the addition of the three new data collection instruments, and the addition of another year of program implementation involving a second cohort of students, the total annual data collection burden (for each year the study is conducted) is estimated at 1,998 hours for an estimated 6,953 respondents and an average of approximately 0.29 hours per respondent. This represents an increase of 1,023 hours in total data collection burden and an average of approximately 0.04 hours per respondent.

16. Plans for Tabulating and Publication of Results

Project Reports

Because of the additional year of data collection from a second cohort of students we will produce an additional report beyond the two we proposed in the original submissions. The two interim and one final report for the project will all focus on assessing the impacts

of the supplemental literacy interventions on various student outcomes and reflecting the research priorities described earlier. Differences in the contents of the three reports will derive from the length of the follow-up period for each cohort of students and the different data that will be available for each.

The first report, due in draft form in October 2006 and in final form in January 2007, will follow students in the first study cohort through the spring of the 2005-2006 school year. At this point, we will have data on student literacy outcomes from the GRADE assessment and Student Follow-up Questionnaire administered in the spring of 2006, and our impact analysis will center on this relatively short-term outcome. The report will also discuss findings from the teacher surveys and consider what we have learned from implementation observations about the quality of implementation of the ERO programs and about the differences between program and control group members in their exposure to literacy skills development activities.

The second report, due in draft form in February 2008 and in final form in May 2008, will analyze the impacts of the interventions on a wider set of outcomes and for both cohorts of students in the study sample. The report will include two years of follow-up for Cohort 1 (through the end of the scheduled 10th grade year) and one year of follow-up for Cohort 2 (through the end of 9th grade). Data for Cohort 2 will include the GRADE Assessment and information from the follow-up survey. In addition, school records data will be available for both cohorts of students including information on attendance, course-taking, and promotion. For Cohort 2, we will have information on student performance on state or district standardized tests for those that administer them in the 10th grade. The second report will also explore the relationship between program effects on literacy and reading comprehension skills (from the GRADE Assessment) and impacts on course taking, promotion and state test scores.

The third and final report, due in draft form in February 2009 and in final form in May 2009, will analyze the impacts of the interventions through three years of follow-up for Cohort 1 (through the end of the scheduled 11th grade year) and two years of follow-up for Cohort 2 (through the end of the scheduled 10th grade year). School records will comprise the primary source of new data for this report. These data will be available for both cohorts of students including information on attendance, course-taking, promotion, and standardized test scores. This report will provide a deeper exploration of the relationship between program effects on literacy and reading comprehension skills (from the GRADE Assessment) and impacts on course taking, promotion and performance on high stakes tests. Finally, this final report will examine factors that are likely to account for the impacts (or lack of impacts) we observe through quasi-experimental methods that build as carefully on the experiment as possible.

Analysis

The first set of analyses would measure the impacts of the supplemental literacy interventions, taken together, on students' reading proficiency, as measured by standardized tests

of reading achievement. Improving students' literacy is the bottom-line goal of these interventions, and as such, it is the primary focus of analyses. We will measure effects not only on overall literacy scores but also on specific dimensions of literacy that are tapped in the ninth-grade reading assessments. We would also examine the impacts of the interventions on other outcomes such as attendance, course-taking and completion, promotion, and persistence in school. Because of the large sample size, we can be most confident about the inferences we draw from these analyses.

The next set of analyses would concern differences in program effects. We would look at the impacts on literacy and other outcomes of each distinct intervention. We would also ascertain whether any differences in the impacts registered by the two interventions are statistically significant – that is, whether one intervention has significantly greater effects than the other. And we would assess whether the effects of the interventions (both collectively and individually) differ for subgroups of students – i.e., students with different characteristics, such as race, socio-economic status, and, perhaps most importantly, prior academic performance. The reduced sample size, along with related considerations of statistical precision, makes us somewhat less confident about the inferences we can draw from our analysis of the impacts of each separate program and our subgroup analyses.

Finally, we will explore the mechanisms through which these interventions affect student outcomes. In particular, we will examine the relationship between variation in impacts and variation across sites in program implementation and in the differences between services received by program and control group members. We will also explore the relationship between impacts on literacy skills and other student outcomes. Because some of these analyses may be non-experimental, they would yield the least definitive conclusions.

These analyses will proceed as described for both cohorts of students. They will include the outcome measures from the Student Follow-Up Questionnaire.

Specific Analyses

As described above, the analyses would proceed at three levels.

First-level analyses: Effects of the interventions, taken together, on all students. We first estimate average impacts for all students across all the 32 participating schools and both intervention models, since these analyses will provide the most statistically robust estimates of the effects of supplemental literacy interventions on student outcomes. While some might question whether the average effect of both interventions across all schools and subgroups is a meaningful and policy-relevant concept, we believe that a major question of this evaluation is whether, on average, these interventions yield meaningful effects on the academic outcomes of struggling ninth-grade readers. We note that although there will be some differences between the two literacy programs to be tested, the basic form of the intervention (i.e., a daily additional period of literacy instruction focused around key components of literacy) will be consistent across both interventions. And we would argue that the target group for the study – ninth-grade students who are two to four grades behind in reading – represents, as a whole, a policy-relevant population, irrespective of possible differences among subgroups.

Our estimates will be obtained from cross-sectional comparisons of students randomly assigned to the program group with students assigned to the control group. For example, we will estimate program impacts on overall literacy, and on key elements of literacy, by comparing program and control group scores on literacy assessments administered in the spring of ninth grade. This analysis would be repeated for the other available outcome measures. This analysis will also be repeated for the next year's achievement data.[7]

To produce valid statistical tests, our experimental program impact estimates must account for the fact that students will be randomized within schools. In order to maximize precision, these estimates will utilize baseline covariates (for example, baseline performance on literacy tests) to reduce unexplained variation in test score outcomes. We can meet this objective by fitting a two-level hierarchical model of the following type to estimate program impacts:

Level 1: Students within Schools

$$Y_{ij} = \beta_{0j} + \beta_{1j} T_{ij} + \beta_{2j} X_{ij} + e_{ij} \quad (1)$$

Level 2: Schools

$$\beta_{0j} = \beta_0 + \mu_{0j} \quad (2)$$

$$\beta_{1j} = \beta_1 + \mu_{1j} \quad (3)$$

$$\beta_{2j} = \beta_0 + \mu_{0j}, \quad (4)$$

Where,

Y_{ij} = literacy achievement for student i at school j;

T_{ij} = 1 if student I is randomly assigned to the treatment group; 0 otherwise, and

X_{ij} = individual student characteristics (e.g., prior achievement) of student i at school j.

In this model,

β_{0j} = average achievement among control group at school j;

β_{1j} = difference between average achievement among treatment and control groups at school j;

β_{2j} = relationship between individual student characteristics (e.g., prior achievement) and student achievement at school j;

β_0 = the average control group outcome across schools in the sample; and

β_1 = the average treatment effect across all of the schools in our sample. This is our primary impact estimate.

[7] It is important to note that, depending on how this analysis is conducted, the examination of effects on tenth-grade achievement may be non-experimental, and therefore, less dependable. Not all students will be promoted to the tenth grade and remain in school long enough to take the achievement test administered in the spring of the tenth grade year. Moreover, the program may have an impact on both promotion rates and persistence in school, thereby having an impact on the composition of students from the treatment and control groups who take this test. As such, comparisons of average test scores might violate the experiment. The only way to generate valid experimental estimates of the effects on student achievement in the tenth grade would be to focus on a dichotomous variable equal to one if the students passed the tenth grade achievement test, and equal to zero in all other cases, including if test scores were missing.

Several points are important to note about this estimator. First, it is based directly on the randomization of students. Therefore, it has the full methodological protection provided by the evaluation's experimental design and represents a reliable estimate of the effect of the supplemental literacy interventions on student outcomes. Second, to ensure that the experiment is maintained, the estimator must be based on data for all of the students who were randomly assigned, not just those who actually participated in the program. Third, this estimator can be used to measure impacts on reading achievement for the different outcome measures described earlier. For example, to measure impacts on the proportion of students who meet a specified reading threshold, one could set Y_{ij} equal to one for students who achieve the threshold and zero for those who do not.[8]

A fourth and final important feature of this estimator has to do with the treatment contrast that it reflects (i.e., the difference in literacy instruction received by students in the program and control groups). In the program group, there may be variation across schools in their implementation of the supplemental literacy interventions. Also, within schools there may be wide variation across program-group students in their exposure to these supplemental services. Furthermore, there will be variation in the literacy instruction that students in the control group receive. To properly interpret the impact, therefore, will require clear documentation of the instructional differences that created it.

Second-level analyses: Effects of each intervention, and effects on subgroups. We will address additional questions pertaining to the effects of each specific intervention and to the effects on various subgroups of students. The analyses that will be conducted in answering these questions preserve the strengths of the experimental design but yield estimates that, because of the smaller sample size, are less precise than estimates derived from the full sample.

Variation in effects by intervention. Our analysis will address three questions: Does Intervention A produce statistically significant effects on literacy and other outcomes? Does Intervention B produce statistically significant effects on the same outcomes? And finally, is there a statistically significant difference between the impacts produced by Intervention A and those produced by Intervention B?

Our approach will be to incorporate into the second level of our model a school-level dichotomous variable indicating which program was in place. In particular, to incorporate differences in effects between two treatments (Treatments A and B) into our model, we would change equation 3 (above) to

$$\beta_{1j} = \gamma_{10} + \gamma_{11} TB_j + \mu_{1j} \qquad (5)$$

where

[8] A more sophisticated logistic formulation of this model might be used instead, but it is unlikely to make much difference unless either a very small proportion or a very large proportion of the students meets the threshold.

TB_j = 1 if school j was randomly assigned to "Treatment B"; 0 otherwise. [9]

In this formulation, γ_{11} yields an unbiased estimate of the difference in program effects between the two interventions.

A further question arises about whether any differences between the impacts achieved by the two interventions are themselves statistically significant. As we indicated in our discussion of minimum detectable effects (B2c), the sample sizes available for this analysis should enable us to detect reliably any differences in program effects larger than .21 standard deviations. While there may be policy-relevant differences that fall within this range, this is below the target effect size for the study.[10]

Variation in effects by student subgroup. A key question relates to variation in program effects across students with different characteristics. An important policy question for this evaluation is the extent to which particular groups of students benefit from the types of supplemental services we will examine. Important subgroups of students include students with different levels of prior achievement, students from different ethnic groups, and students with different instructional statuses (e.g., English language learners, special education students).

The research team would use information recorded on the baseline Student Background Questionnaire pertaining to students' characteristics before random assignment in order to identify sub-samples of students for these analyses. The team would then estimate the two-level model described above for each subgroup in order to identify effects for these particular students. We would also attempt to ascertain the extent to which these effects differ across particular groups of students. This can be accomplished by adding an interaction term between the experimental variable and particular pre-random assignment characteristics[11] into the system of equations described above. For example, a question of interest is whether or not supplemental literacy interventions have the same effect for very poor children as for other children. In order to explore this question, we would alter our impact model in the following manner.

Level 1: Students within Schools

$$Y_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \beta_{2j}X_{ij} + \beta_{3j}R_{ij} * T_{ij} + e_{ij} \qquad (5)$$

Level 2: Schools

[9] Alternatively, because schools were randomly assigned to receive Treatment A or B, we would expect the control group means from Treatment A schools to be identical to those in Treatment B schools. Therefore, we could also estimate the difference between the impacts of the treatments by comparing the treatment group means from Treatment A to those in Treatment B.

[10] We could use a similar approach to test the effects of teachers' background characteristics and preparation, measured on the teacher survey, on student outcomes. While schools have been randomly assigned to interventions, however, teachers have not, so that this analysis should be considered more exploratory than definitive. Moreover, since there will be only 32 intervention class teachers in all and only 16 teachers per intervention, the small samples will reduce our confidence in the conclusions we might draw.

[11] This can also be accomplished by estimating the effects among particular subgroups of students in separate analyses, and then comparing the impacts from the two different samples. The distinction between these split-sample techniques and those using interactions is that the latter assumes that the relationships between other covariates and the outcome variables are the same across the sub-groups in the sample. This assumption may affect statistical precision, but it does not affect the validity of the estimate one way or another.

$$\beta_{0j} = \beta_0 + \mu_{0j} \quad (6)$$

$$\beta_{1j} = \beta_1 + \mu_{1j} \quad (7)$$

$$\beta_{2j} = \beta_2 + \mu_{0j}, \quad (8)$$

$$\beta_{3j} = \beta_3 + \mu_{0j}, \quad (9)$$

where

R_{ij} = 1 if student i at school j is poor (i.e., eligible for free or reduced-price lunch), 0 otherwise.

In this formulation,

β_1 = the average impact of the interventions among students who are not poor, and

β_3 = the difference between the effect of supplemental literacy interventions on poor students compared to their counterparts who are not poor.[12]

17. OMB Expiration Date

All data collection instruments will include the OMB expiration date.

18. Exceptions to Certification Statement

No exceptions are requested.

[12] It should be noted here that accounting for the nesting of students within schools avoids confusing differences in effects across sub-groups of students with differences in the effectiveness of the various sites.

PART B: COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

This is an evaluation of literacy programs for 9th grade students who read between two and four years below grade level to be conducted in 32 sites to be selected by the Department of Education's Office of Vocational and Adult Education. The sites were selected through a grant competition done by OVAE. Upon selection, MDRC randomly assigned each high school to one of the two interventions so that each district will have two high schools conducting each intervention.[13]

Since sample selection and random assignment to the program or control group will be done at the start of 9th grade, the respondent universe for this evaluation consists of all 9th grade students who read between the 7th and 5th grade level at the selected sites.

The research design calls for 100 students – 50 in the control group and 50 in the program group – in each site. About 150 students will be sent the consent form and Student Background Questionnaire per site. If there are more than 150 eligible students, 150 will be selected using a probability sample. We expect about 20% to not return the form, resulting in 120 potential participants. Again, the 100 students in the sample will be selected using a probability sample. In terms of the total sample 4800 will be sent the forms; we expect 3840 to return the forms; there will be 3200 in the study although all selection will be done at the school level.

Although this is the ideal design, we assume that operating the literacy interventions with fewer than 50 students in a high school may be inefficient, and that program openings resulting from attrition from the treatment group could create opportunities for other eligible students to enroll, especially if that attrition occurs early. If our assumptions are correct, and if the programs can accommodate new entrants after the start of the year or during the second semester, we would want to ensure that the process for adding additional students to maintain program capacity is fair and transparent and that students from the control group continue to be excluded from the literacy classes. We could accomplish these goals by building two additional features into the random assignment process. First, we could randomly assign more than 50 students to the literacy interventions to allow for some attrition. Second, depending on the number of students in the initial pool, we could randomly assign students to a waiting list of about 10 students who would not be included in the research sample but would be allowed to enroll in the literacy classes if openings occurred. We would discuss these options with administrators at the participating schools.

[13] In each high school there will be one teacher implementing the intervention (i.e., teaching four supplemental literacy classes of 12-15 students each and assisting with research needs). A second teacher from each school will attend the professional development for the intervention, serving as a back-up in case the original teacher needs to leave the program during the school year.

2. Information Collection Procedures

a. Statistical methodology and stratification

As described above the study sample will be a random probability sample of the universe of students who returned forms.

b. Estimation Procedures / Analysis Methods

See section A.16 for a full description of the analysis plan.

c. Degree of Accuracy Needed

General considerations and the minimum detectable effect size metric. To ensure that the evaluation will produce reliable and valued findings, the sample size must be large enough to enable the study to answer the study's central questions and to measure program effects that are large enough to be both meaningful in students' lives and relevant to policy debates about the efficacy of supplemental literacy interventions.

In the discussion that follows, we report precision as "minimum detectable effect sizes."[14] This metric, which is used widely for measuring the impacts of educational programs, is defined

[14] We define a minimum detectable effect as the smallest true program impact that would have an 80 percent chance of being detected (have 80 percent power) using a one-tail hypothesis test at the 0.05 level of statistical significance. We use a one-tail test because the central policy issue to be addressed by the evaluation is whether the program *improves* student reading performance. Based on calculations derived from Bloom (2003), we estimate minimum detectable effect size as follows:

$$MDES = 2.8 * \sqrt{\frac{\sigma_y^2(1-R^2)}{P(1-P)(n)(J)(\sigma_y^2+\tau_y^2)} + \frac{\omega^2}{J(\sigma_y^2+\tau_y^2)}},$$

where:

$\sigma_y^2 =$ the (within site) variance of the outcome in question (assumed to be 1; however, by definition of effect size metric, does not affect the minimum detectable effect size);

$R^2 =$ the explanatory power of the impact regression, i.e., the proportion of the variance in y explained by the experiment. In order determine the appropriate r-square, MDRC regresses 9th grade SAT-9 achievement on 8th grade scores for high school students in the Houston school district in 2002. The regression produced an r-square value of .69, which we used in our effect size calculations);

$P =$ the proportion of students randomly assigned to the treatment group (assumed to be .5 based on balanced random assignment design);

$n =$ the number of students in each site);

$J =$ the number of sites in the study.

$\tau_y^2 =$ the cross site variance in the mean value of the outcome measure y (calculated based on an assumption that the intraclass correlation $\frac{\tau^2}{\tau^2+\sigma^2} = .07$, an assumption based on MDRC's analysis of achievement data across all comprehensive non-exclusive high schools in the Houston school district;

in terms of the underlying population standard deviation of student achievement. For example, a minimum detectable effect size of .25 indicates that an impact estimator can reliably detect a program-induced increase in student achievement that is equal to or greater than .25 standard deviations of the existing student distribution.

No absolute standard exists as to what represents a large versus a small effect size. A meta-analysis of treatment effectiveness studies sheds some light on this issue (Lipsey, 1990). This study found that, out of 102 studies, most of which were from education research, the bottom third of the distribution of impacts ranged from about 0 to .32 effect size, the middle third of impacts ranged from .33 to .50, and the top third of impacts ranged from .56 to 1.26. For this discussion we are taking .25 as a reasonable target for detecting fairly small effects.

The proposed sample and its adequacy for answering the study's key questions. Table 6 shows the sample sizes and minimum detectable effects for various configurations of site and student subgroups. The second column shows sample sizes in the ideal case that follow-up data are available for all students in the sample.[15] The basic design calls for a total sample of 34 schools. Within each high school, 100 students would be randomly assigned to program and control groups. As shown in the second line of Table 4 this would yield a total sample of 3400 students. The remaining rows in the exhibit provide insight into potential subgroups that may be of interest (e.g., students of different ethnic backgrounds, genders, or levels of prior achievement). The third column presents our assessment of how minimum detectable effect sizes for estimates of impacts on average achievement scores would vary with different sample sizes and configurations of sites and student subgroups.

ω^2 = the cross site variance in the true impact of the program (assumed to be 0 based on the specification that the impacts would be estimated as "fixed effects" rather and random or varying effects across sites). This "fixed effect" assumption was specified by the U.S. Department of Education.

This estimate of the minimum detectable effect size accounts for both within site and across site variation in the outcome in question. In this equation, the expression $\frac{\sigma_y^2(1-R^2)}{P(1-P)(n)(J)(\sigma_y^2+\tau_y^2)}$ represents the error contributed by the estimation error within each of the sites we are studying.

[15] We discuss below the effects of attrition from the sample. We anticipate, however, that follow-up data from administrative records will be available for the large majority of students participating in this study.

Table 6: Sample Sizes and Minimum Detectable Effects

	Sample Size	Minimum Detectable Effect	Sample Size if 80 Percent Response rate	Minimum Detectable effect if 80 Percent Response Rate
All Sites, All Students	3400	.05	2720	.06
All sites, 25 Percent subgroups	850	.10	680	.12
17 sites, all students	1700	.07	1360	.08
17 sites, 25 Percent Subgroups	425	.15	340	.16
Differences in the effects of interventions, all students	3400	.07	2720	.08
Differences in the effects of interventions, 25 Percent subgroups	850	.12	640	.13

Do the two interventions, considered together, have a statistically significant effect on student outcomes? Our first question is whether the sample is large enough to detect the effects of the two supplemental literacy interventions, taken together. Our statistical power calculations suggest that, with respect to this question, we can be confident that, with the proposed number of students and sites, the evaluation will effectively detect as statistically significant any meaningful changes in test scores and other student outcomes. In particular, the first row in Table 4 indicates that with a sample of 34 sites and 3400 students, we can be confident that the evaluation will detect impacts on test scores as small as .05 standard deviations. This is equivalent to just under 2 Normal Curve Equivalent (NCE) points on a nationally norm-referenced achievement test and translates into the difference between the 25th and the 27th percentile.

The estimated minimum detectable effects sizes for the full sample of 34 sites may be considered small, and we can conclude that if policy-relevant differences are created by the intervention, the evaluation will detect them.

How do the interventions affect different subgroups of students? In addition to questions regarding effects for the full sample of students in the study, the research design must be adequate for detecting effects among students with particular characteristics. The most logical strategy for answering this question is to estimate program impacts among sub-samples of students with particular characteristics, such as race, gender, socioeconomic status, and, perhaps most importantly, prior achievement. Columns four and five in Table 4 present sample sizes and the estimated minimum detectable effect sizes assuming a subgroup 25 percent. As is the case for the full sample, the effect sizes presented in the table suggest that the study design is sufficient to detect relatively modest program effects among student subgroups. Even for a subgroup consisting of 25 percent of the sample, the estimates in the table indicate that the sample from all

34 schools is sufficient to detect an effect of .10 standard deviations. This is only a slightly larger than the minimum detectable effects for the full sample.

Does each separate intervention have a significant effect on outcomes? The two treatments will be randomly assigned across the 34 schools in the sample, so that 16 schools will implement one treatment, and the remaining 17 schools will implement the other. The proposed sample size and configuration also needs to be sufficient to assess the effects of each treatment separately. The estimated minimum detectable effect size for a sample of 17 schools (the third line of Table 4) is .07 standard deviations, again smaller than the target effect size. So with the proposed sample size, we are likely to detect most policy relevant effects of each specific intervention.

The sample sizes for subgroup analysis within the each treatment are considerably smaller and result in a minimum detectable effect of .15 (line four).

Are differences in the effects of the two interventions statistically significant? Reliable estimates of the differences in program impacts across the two treatments would be highly relevant to policy and practices. We build on the cluster random assignment design and the analysis conducted in Bloom (2003) in order to estimate the minimum detectable effect sizes for this comparison. Line 5 of Table 4 reports estimated minimum detectable effect of the difference in impacts for this cluster random assignment study. These estimates indicate that the design is sufficient to detect the presence of meaningful differences in program impacts. Assuming 16 sites per treatment, the minimum detectable effects are .07, lower than the target effect size of .25. This suggests that if meaningful, systematic differences between the two treatments exist, the study is likely to detect them.[16]

Response rates and attrition. We will do everything possible to ensure that data are collected for as many of the respondents as possible. Nevertheless, there are likely to be some limits on our ability to collect data. For example, we can probably collect achievement test data from all students who remain in the same school district, even if they change schools. However, we most likely will be unable to collect follow-up achievement data on students who move out of the district they were in at the point of random assignment.

Reduced sample sizes resulting from attrition affect the ability of the study to reliably detect impacts. Given that random assignment and the initial follow-up occur within the same year, it is likely that attrition from the sample will be limited. As a conservative estimate, the fourth column of Table 4 shows sample sizes assuming follow-up data are available for 80 percent of the students in the sample.[17] The minimum detectable effect sizes – given in the last column of Table 4 - do not differ much from those calculated assuming 100 percent response rates, suggesting that, even with some attrition, the evaluation will be able to provide reliable answers to questions regarding the overall treatment effect.

[16] On the other hand, it is important to remember that if one treatment had an effect of .2, and the other had an effect of .4 – a difference of .2 rather than .22 standard deviations -- this sample might not be sufficient to reliably distinguish between the two effect sizes.

[17] Note that the panel of the table assumes that there is no non-response bias in data availability.

3. Methods to Maximize Response Rates

Only students who return the parental consent forms can participate in the evaluation, so it will be critical to get a high rate of return of these forms.[18] Our experience working in high schools both as practitioners and researchers suggests that this is likely to be a difficult and time-consuming process. Getting a high rate of return will require sending a strong message to the students and parents about the valuable opportunity presented by the supplemental literacy classes and the importance of completing and returning the forms as quickly as possible. It may also be necessary to contact students and parents by phone or through the mail. Study liaisons stationed at the schools will develop procedures for contacting students and their parents during both the school year and the summer to have them return the forms.

MDRC and AIR will work with the high schools to most efficiently administer the GRADE. Students who were absent the day of the test will be followed up and given another chance to complete it. Students who have changed schools will be followed and we will try to arrange completion of the test for them as well.

4. Tests of Procedures to be Undertaken

The items in the Student Background Questionnaire are taken almost directly from the very successful NAEP 8th grade baseline survey. **The Student Follow-Up Questionnaire also includes many of the same items from the NAEP survey as well as items adapted from the National Education Longitudinal Survey (NELS), the ALD Out-of-School Literacy Practice and ALD In-School Literacy Practices student surveys used in the NICHD-funded Study of Social and Cultural Influence on Adolescent Literacy Development, and the Media Research Questionnaire used by Hobbs et al. to study media literacy.** Items from all the teacher surveys are drawn from teacher surveys from the National Longitudinal Evaluation of Comprehensive School Reform, the Evaluation of the National School District and Network Grants Program (Gates Foundation), the Alabama Reading Initiative, the Teacher Belief System (Benjamin, 2003), and the Survey of Instructional Practices and Content for English Language Arts and Reading (Survey of Enacted Curriculum, 2004). We will conduct a small number of pre-tests to determine the exact amount of time it takes to complete these instruments and to be sure that all the questions are clear to the respondents.

The GRADE is a widely used group-administered, paper-and-pencil test. Because subtests can be administered separately, the test can be divided across two or more sessions to accommodate schools' class schedules. The directions for administering the GRADE are easy to follow, so school staff and the research team can readily accomplish this task. Tests can be easily scored using software provided by the developer.

[18] We recognize that students who complete and return all the forms may be better organized or more motivated than those who do not, and that the generalizability of the study results may be somewhat limited if the supplemental literacy interventions turn out to serve a more select group of struggling readers than would typically be found in low-performing high schools. Nonetheless, we view completion of these forms – especially the informed consent form – as an important condition of students' participation in the evaluation.

5. Individuals Consulted on Statistical Aspects of Design

Dr. James Kemple, MDRC

Dr. Howard Bloom, MDRC

Dr. Jason Snipes, MDRC

In addition to the above, an evaluation advisory panel has provided substantial input on the study design and data collection plan. The advisory panel members represent a number of the nation's leading researchers on adolescent literacy and evaluation design. The panel includes:

- Dr. Donna Alvermann, University of Georgia
- Dr. Donald Compton, Vanderbilt University
- Dr. Robinson Hollister, Swarthmore College
- Dr. Mark Lipsey, Vanderbilt University
- Dr. Robert Meyer, University of Wisconsin
- Dr. Christopher Schatschneider, Florida State University
- Dr. Timothy Shanahan, University of Illinois at Chicago
- Dr. Catherine Snow, Harvard University

REFERENCES

Alliance for Excellent Education. (2004). *Effective literacy instruction for adolescents.* National Reading Conference.

Biancarosa, G. & Snow, C.E. (2004). *Reading Next—A Vision for Action and Research in Middle and High School Literacy: A Report from Carnegie Corporation of New York.* Washington, DC: Alliance for Excellent Education.

Bloom, Howard. (2003). *Sample Design for an Evaluation of the Reading First Program*. MDRC Working Papers on Research Methodology. New York: MDRC.

Caldwell, J. & Leslie, L. (2004). Does proficiency in middle school reading assure proficiency in high school reading? The possible role of think-alouds. *Journal of Adolescent and Adult Literacy, 47*, 324-335.

Grigg, W. S., Daane, M. C., Jin, Y., & Campbell, J. R. (2003). *The Nation's Report Card: Reading 2002.* Washington, DC: National Center for Education Statistics.

Group Reading Assessment Diagnostic Evaluation. (2001). Circle Pines, MN: American Guidance Service, Inc.

Guthrie, J.T. & Alvermann, D.A. (1999). *Engaged Reading: Processes, Practices and Policy Implications*. Teachers College Press: New York.

Guthrie, J. T. (2002). *Proceedings of Adolescent Literacy – Research Informing Practice: A Series of Workshops*. Washington, DC: National Institute for Literacy.

Kamil, M. (2004). *Adolescents and Literacy: Reading for the 21st Century.* Washington, DC: Alliance for Excellent Education.

Kemple, J., & Herlihy, C. (2004). *The Talent Development High School.* New York: MDRC.

Lipsey, Mark. (1990). *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: Sage Publications.

Moore, D. W., Bean, T. W., Birdyshaw, D., & Rycik, J. A. (1999). *Adolescent literacy: A position statement*. Newark, DE: International Reading Association.

National Assessment Governing Board. (2003). *Background information framework for the National Assessment of Educational Progress.* Washington, DC: Author.

National Center for Education Statistics. (2004). *The condition of education 2004.* Washington, DC: National Center for Education Statistics.

National Reading Panel (NRP). (2002). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Washington, DC: National Institute of Child Health and Human Development.

National Research Council (NRC). (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.

Reardon, S., & Galindo, C. (2002). *Do high-stakes tests affect students' decisions to drop out of school? Evidence from NELS*. Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Rycik, J. A., & Irvin, J. L., (2001). *What adolescents deserve: A commitment to students' literacy learning.* Newark, DE: International Reading Association.

Wigfield, Allan. (2004). Motivation for Reading During the Early Adolescent and Adolescent Years. In D. Alvermann and D. Strickland (Eds.), *Bridging the Literacy Achievement Gap: Grades 4-12* (pp. 43-55). New York: Teachers College Press.